

PUBLICATION NUMBER : 10136327  
PUBLICATION DATE : 22-05-98

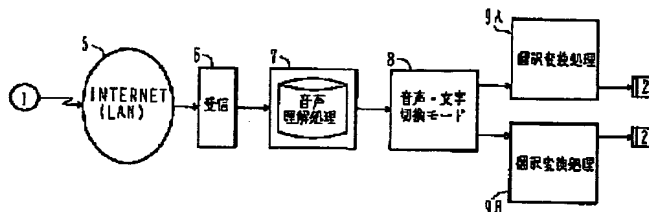
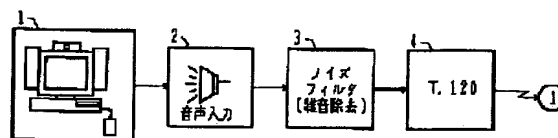
APPLICATION DATE : 25-10-96  
APPLICATION NUMBER : 08283462

APPLICANT : MEIDENSHA CORP;

INVENTOR : TAKAHASHI KAZUHIKO;

INT.CL. : H04N 7/15 G06F 17/28 G10L 3/00  
G10L 3/00 G10L 3/00 H04M 3/42  
H04M 3/56

TITLE : DESK TOP CONFERENCE SYSTEM



ABSTRACT : PROBLEM TO BE SOLVED: To provide the desk top conference system in which smooth communication is attained even when participants of a conference speak different languages.

SOLUTION: Speeches of conference participants are received via a communication system (1-6) and a speech understanding device 7 obtains character data of a language understandable by a recipient and corresponding to the speech signal of each participant, a changeover made section 8 sets whether the recipient receives the character data in characters or in voice in a language understandable by the recipient. A translation conversion processing section 9A or 9B translates the character data into character data or speech data in the language depending on the setting, the translated character or speech data are converted into a text or a voice in the language, displayed on the display device together with a received video image in terms of captions or recorded and displayed on an editor and synthesized by a speech synthesis device and outputted in voice.

COPYRIGHT: (C)1998,JPO

(11)特許出願公開番号

特開平10-136327

(43)公開日 平成10年(1998)5月22日

FI

H04N 7/15

G 1 0 L 3/00

R

5 5 1 C

5 6 1 G

P

審査請求 未請求 請求項の数 1 OL (全 11 頁) 最終頁に続く

(71)出願人 000006105

株式会社明電舎

東京都品川区大崎2丁目1番17号

(72)発明者 高橋 和彦

東京都品川区大崎2丁目1番17号 株式会社明電舎内

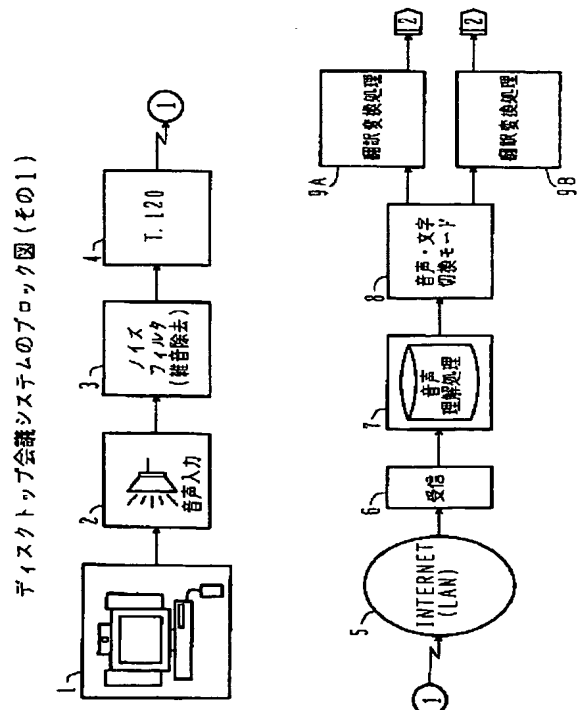
(74) 代理人 弁理士 志賀 富士弥 (外1名)

(54) 【発明の名称】      ディスクトップ会議システム

(57) 【要約】

【課題】 映像と音声の通信で互いに言語の異なる人が  
ディストロップ会議を行うときには、言語の違いにより  
円滑なコミュニケーションが難しい。

【解決手段】 会議出席者の音声通信システムを介して受信し（１～６）、この出席者の音声信号に対応した受信者が理解できる言語の文字データを音声理解装置７に得、この文字データを受信者が理解できる言語の文字または音声で受け取るかを切換モード部８で設定し、この設定に応じて翻訳変換処理部９Ａまたは９Ｂにより文字データを該言語の文字データ又は音声データに翻訳し、翻訳した文字又は音声データを該言語の文章又は音声に変換して受信映像と共にディスプレイに字幕表示又はエディタに記録表示、さらには音声合成装置で合成して音声で出力する。



## 【特許請求の範囲】

【請求項1】 会議出席者の音声を通信用システムを介して受信する音声信号受信手段と、  
前記音声信号のA/D変換データを周波数スペクトルによって大まかに分析し、特徴抽出によって音声信号の時系列の特徴パラメータを音素単位に抽出し、各音素単位の認識に続く単語の認識、及び構文認識と意味認識により前記出席者の音声信号に対応した文字データを得る音声理解装置と、  
前記文字データを受信者が理解できる言語の文字データ又は音声データに翻訳する翻訳変換処理部と、  
前記翻訳した文字又は音声データを前記言語の文章又は音声に変換する文字・音声変換部と、  
前記変換部からの文章を受信映像と共にディスプレイに表示又はエディタに記録表示する出力切換部と、  
前記変換部からの音声を音声合成装置で合成して出力する音声出力装置と、を備えたことを特徴とするディスクトップ会議システム。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、海外諸国との会議室相互間を映像及び音声でコミュニケーションを得るためのディスクトップ（テレビ）会議システムに関する。

## 【0002】

【従来の技術】図14は、ディスクトップ会議システムの実現形態例を示す。このシステムには、特定用途向けのディスクトップ会議システムがあり、ディスクトップ会議システムの機能を重視して画質や音質にこだわる。例えば、消費者金融の無人契約機や遠隔監視システム、コンサルティングなどでの利用が拡大しつつある。

【0003】また、オフィスで使うパソコン（PC）の標準機能を目指したディスクトップ会議システムもある。これはディスクトップ会議よりもデータ会議の機能を重視している。

【0004】また、ディスクトップ会議やデータ会議の機能を必要最低限に割り切ったディスクトップ会議システムもある。代表例には米国White Pine Software, Inc. のディスクトップ会議ソフトウェアがある。

【0005】図15は、他のディスクトップ会議システムとして、ActiveX Conferencing のアーキテクチャを示し、米国の"Professional Developers Conference"で発表されたものである。

【0006】このシステムは、インターネット上で音声や動画、データをリアルタイムでやり取りするためのプラットフォームと位置づける。基本機能は、一連の国際標準（ITU-T勧告）を採用している。例えば、音声コーデック（圧縮／伸長）はG. 723、動画コーデックはH. 263、多地点データ会議用プロトコルは

T. 120、呼制御はH. 245などにそれぞれ準拠している。音声データや動画データをリアルタイム伝送するためのプロトコルは、インターネットの標準化組織であるIETFが決めたRTP（Real Time Protocol）を使用している。

【0007】図16は、T. 120プロトコルの構成を示し、大きくには上位プロトコルと下位プロトコルの2階層に分かれる。上位プロトコルは、機能別にホワイトボード共有用のT. 126やファイル転送用のT. 127などがある。ディスクトップ会議制御用のT. 130は、1997年にも運用される。アプリケーション共有は、T. 120の対象外である。

【0008】下位プロトコルには多地点データ会議制御用のT. 122/T. 125/T. 124と、伝送用のT. 123がある。T. 120ではさまざまなT. 123を用意することで、ネットワークの違いを吸収するというものである。

【0009】図17は、T. 120プロトコルを実装したディスクトップ会議システムと多地点接続装置（MCU）を示す。MCUは、T. 120の下位プロトコルだけを実装し、上位プロトコルはディスクトップ会議システム同士がエンド・ツー・エンドでやり取りする。1996年3月のIMTCの相互接続試験では、下位プロトコルだけが確認されている。すなわち、T. 120準拠のMCUとディスクトップ会議システムは、異なるメーカーの製品同士を相互接続できる。

【0010】図18は、H. 323準拠のディスクトップ会議システムのアーキテクチャを示す。H. 323の仕様は、動画コーデック、音声コーデック、システム制御、データの packets 化や同期などに関する勧告からなる。動画コーデック用のH. 261は、ISDN利用のディスクトップ会議システム（H. 320）で標準化された方式である。H. 263は、既存のアナログ電話回線を利用するディスクトップ会議システム（H. 324）で標準化された方式である。H. 263はH. 261に比べて符号化効率が2～3倍と優れている。

【0011】音声コーデックには、64Kbit/secのG. 711、帯域幅7KHzの高品質音声を64Kbit/secあるいは5.3Kbit/secのG. 723、16Kbit/secのG. 728、8Kbit/secのG. 729がある。

【0012】以上のディスクトップ会議システム構成の他に、多地点接続機能をWWWサーバに統合したもの、ディスクトップ会議システムと企業内情報システムを統合したものなどがある。

## 【0013】

【発明が解決しようとする課題】従来のディスクトップ会議システムを利用して海外諸国とコミュニケーションを行う場合、以下のようなトラブルが発生し、円滑なコミュニケーションが取れない。

【0014】(1)例えば、英語は英語で、日本語は日本語でしかコミュニケーションが取れない。

【0015】(2)英語を日本語に翻訳し、又は日本語を英語に翻訳し、かつ字幕あるいは音声で出力することができない。

【0016】(3)翻訳モードがない。

【0017】(4)コミュニケーション中、聞き間違えたり聞き逃したりする。

【0018】本発明の目的は、会議の出席者が互いに異なる言語を話す場合にも円滑なコミュニケーションを得ることができるディストップ会議システムを提供することにある。

【0019】

【課題を解決するための手段】本発明は、コミュニケーション上のギャップを解決するために、人の発する音声(英語や日本語)を音声認識装置でリアルタイムで認識し、これを翻訳装置で翻訳(日本語→英語、英語→日本語など)し、この翻訳した文の音声出力や字幕表示さらには会議議事録として出力するもので、以下の構成、会議出席者の音声を通信システムを介して受信する音声信号受信手段と、前記音声信号のA/D変換データを周波数スペクトルによって大まかに分析し、特徴抽出によって音声信号の時系列の特徴パラメータを音素単位に抽出し、各音素単位の認識に続く単語の認識、及び構文認識と意味認識により前記出席者の音声信号に対応した文字データを得る音声理解装置と、前記文字データを受信者が理解できる言語の文字データ又は音声データに翻訳する翻訳変換処理部と、前記翻訳した文字又は音声データを前記言語の文章又は音声に変換する文字・音声変換部と、前記変換部からの文章を受信映像と共にディスプレイに表示又はエディタに記録表示する出力切換部と、前記変換部からの音声を音声合成装置で合成して出力する音声出力装置と、を備えたことを特徴とする。

【0020】

【発明の実施の形態】図1及び図2は、本発明の実施形態を示すディストップ会議システムのブロック図であり、図3～図9に各部の処理過程を同じ符号で対応付けて示す。なお、以下の説明は送信側から受信側への一方の音声系統のみを示し、逆方向になる音声情報及び別系統になる映像情報の送受信系統は省略している。

【0021】ディストップ会議の出席者がパソコン1のマイクロフォン2から音声を入力したとき、この音声信号から音声に伴う雑音をノイズフィルタ3により除去する。

【0022】通信制御装置4は、多地点データ会議用プロトコル(T. 120)によってインターネット5に接続し、雑音除去した音声信号を会議相手の受信装置6に伝送する。これら音声入力処理と通信制御処理は図3に示す。

【0023】受信装置6で受信した音声情報は、ニュー

ロコンピュータシステムになる音声理解処理装置7により音声理解とその文字変換を行う。これら処理は、図4及び図5の手順になり、以下の手順により実現される。

【0024】(a)入力された音声信号をA/D変換し、音声分析により音声信号の周波数スペクトルによって大まかに分析する。

【0025】(b)分析した音声信号の特徴パラメータを抽出して時系列データに変換する。

【0026】(c)時系列データを音声単位にセグメンテーション化を行う。

【0027】(d)時系列データから外国語(音声入力装置2で入力した会議出席者の言語)の音声認識を行う。この認識には、音声標準パターンとの比較後、音素認識を行って音素系列を作成し、音素系列について単語辞書を参照した単語認識を行い、認識された単語が構文的に誤りであるかないかを分析する構文解析を行い、誤りならば単語認識の再検証をし、構文に誤りがなければ、意味的に妥当であるか否かを調べる意味解析を行い、妥当な結果が得られなければ単語認識まで逆上って再検証を行う。

【0028】現在の音声認識技術は、音の歪みや雑音を除去し、不特定話者による音声の認識が可能であり、10万単語の大語彙の実現により、ボタン操作や語彙数の制限による違和感を払拭でき、利用者は日常使用する言葉/単語で字ぜんな音声入力が可能となっている。また、認識単語は、テキストデータにより登録でき、かつユーザ自身が使用したい単語を自在に、しかも簡単・スピーディに設定することが可能である。さらに、単語の追加/削除/変更もスムーズに行うことが可能である。

【0029】図1に戻って、モード切換部8は、音声認識されたデータの出力モードを会議出席者の設定により切換える。翻訳変換処理部9Aは、出力モードが文字モードにされたときに起動され、音声認識されたデータを受信者が理解できる言語の文字に変換しかつ字幕で表示するための翻訳変換処理を行う。翻訳変換処理部9Bは、出力モードが音声モードにされたときに起動され、音声認識されたデータを受信者が理解できる言語の音声に変換しかつ音声で出力するための翻訳変換処理を行う。

【0030】これら翻訳変換処理を図5～図8に示す。モード切換後、音声単語列を知識ベース(学習機能付き単語辞書)へ照合して受信者が理解できる言語の単語を認識する(図5)。これら認識した単語の組み合わせとした構文の解析を知識ベース(学習機能付き文法辞書)へ照合して受信者が理解できる言語の構文を認識する(図6)。これら認識した構文について、文の意味の解析を知識ベース(学習機能付き意味辞書)へ照合して受信者が理解できる言語の文のパターン(文字列)を認識する(図7)。認識不可なら再度に構文の解析に戻る。

【0031】単語と構文と文の意味の認識後、翻訳変換

プロセスに入る(図8)。翻訳が不可なら再度に文の意味解析に戻り、ファジィ推論によって翻訳可能な文にまとめ、翻訳を実行する。また、翻訳処理の実行中は、訳語の単語・文法・構文を知識ベース(学習機能付き単語辞書、文法辞書、パターン認識)に照合し、訳語を序列化する。この後、文法チェック機能によって受信者が理解できる言語の文または文法を修正・生成する。

【0032】生成した文は、切換モードが音声か文かによって、図2の文字変換部10A又は音声変換部10Bにより文字データ又は音声データに変換する(図9)。このうち、文字データは、出力切換部11Aによりディスプレイ12に文字表示、又はエディタ13に記録表示する。

【0033】音声データは、音声合成装置11Bにより受信者が理解できる言語の音声合成を行い、文章の内容を音声出力装置14に音声として出力する。

【0034】以上までの構成と処理により、会議出席者が互いに異なる言語でディスクトップ会議を行う場合にも、発信者は自身が理解できる言語で発言し、この発言を受信者が理解できる言語に翻訳して相手のディスプレイに表示や音声出力、さらには議事録として記録することができ、海外諸国とのコミュニケーションを円滑に行うことができる。

【0035】以上までのシステム構成に必要なインタフェースは、図10に示すようになる。

【0036】なお、図1における翻訳変換処理9A、9Bの翻訳処理は、例えば、トランスファー方式により実現される。このトランスファー方式は、図11に示すように、言語列の中間表現を持ち、中間表現で言語の変換を行う。例えば、英語→日本語や日本語→英語といったように翻訳する言語に中間表現が作られる。解析の手法としては、日本語に格文法、英語には拡張遷移文法がよく使用される。

【0037】図12は、日本語から英語に翻訳するときのシステム構成を示す。日本文の構文解析と意味解析を行い、その意味を表す日本語の中間表現を得る。その次に、日本語の中間表現を英語の中間表現に変換し、最後に英語の中間表現から英文を生成する。

【0038】図13は、トランスファー方式の他に、元の単語を目的の言語の単語に置き換えてその順序を並べ替えるダイレクト方式と、元の言語を言語の種類に依存しない中間表現に変換してその中間表現から目的とする言語を生成するピボット方式と、これら方式にトランスファー方式の特徴を組み合わせたシステムを示し、より複雑な翻訳処理を可能にする。

【0039】

【発明の効果】以上のとおり、本発明によれば、会議出席者が互いに異なる言語でディスクトップ会議を行う場合にも受信者が理解できる言語に翻訳して相手のディスプレイに表示や音声出力、さらには議事録として記録す

ることができ、海外諸国とのコミュニケーションを円滑に行うことができる。具体的には、以下の効果がある。

【0040】(1) 翻訳モードの切換えにより、例えば日本語から英語に翻訳しかつ英語の字幕及び音声で受信できる。

【0041】(2) 会議に議事録を残すことができる。

【0042】(3) 言語間の軋轢を回避することができる。

【0043】(4) 音声の他に、字幕表示によって聞き取りにくいところや聞き逃し、聞き間違えなどを回避できる。

【図面の簡単な説明】

【図1】本発明の実施形態を示すディスクトップ会議システムのブロック図(その1)。

【図2】本発明の実施形態を示すディスクトップ会議システムのブロック図(その2)。

【図3】図1または図2における各部の処理過程(その1)。

【図4】図1または図2における各部の処理過程(その2)。

【図5】図1または図2における各部の処理過程(その3)。

【図6】図1または図2における各部の処理過程(その4)。

【図7】図1または図2における各部の処理過程(その5)。

【図8】図1または図2における各部の処理過程(その6)。

【図9】図1または図2における各部の処理過程(その7)。

【図10】本発明における訳語機能付ディスクトップ会議システムのインタフェース例。

【図11】実施形態における翻訳のためのトランスファー方式の構成。

【図12】図11におけるトランスファー方式による翻訳システム構成図。

【図13】翻訳のための3つの変換方式の組み合わせ。

【図14】TV会議システムの実現形態。

【図15】ActiveX Conferencingのアーキテクチャ。

【図16】T.120プロトコルの構成。

【図17】T.120プロトコルを実装したディスクトップ会議システムとMCU。

【図18】H.323準拠のテレビ会議システムのアーキテクチャ。

【符号の説明】

- 1…パソコン
- 2…マイクロフォン
- 3…ノイズフィルタ
- 4…通信制御装置

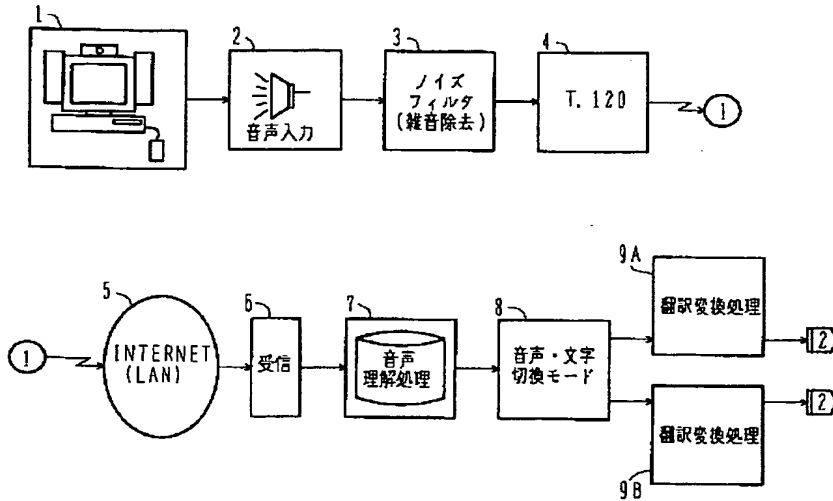
5…インターネット  
6…受信装置  
7…音声理解装置  
8…モード切換部  
9A、9B…翻訳変換処理部  
10A…文字変換部

10B…音声変換部  
11A…出力切換部  
11B…音声合成装置  
12…ディスプレイ  
13…エディタ  
14…音声出力装置

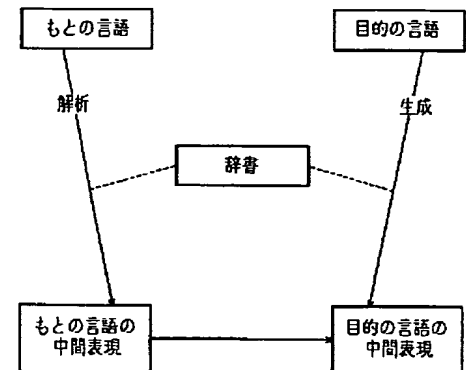
【図1】

【図11】

デスクトップ会議システムのブロック図(その1)



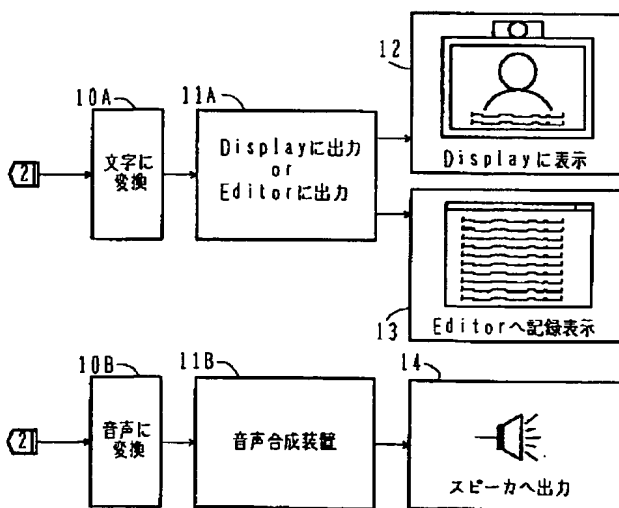
トランスファー方式の構成



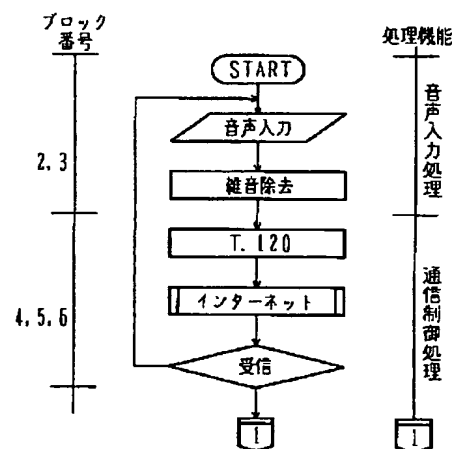
【図2】

【図3】

デスクトップ会議システムのブロック図(その2)

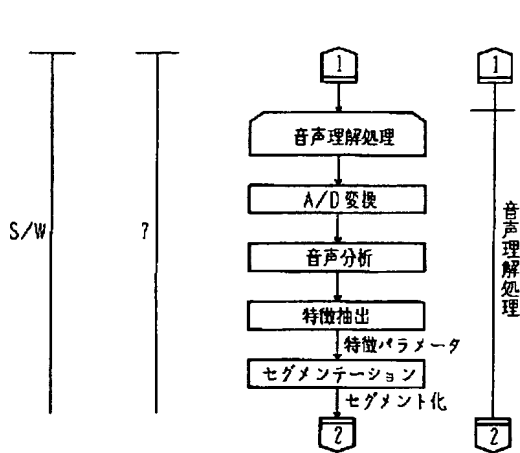


処理過程(その1)



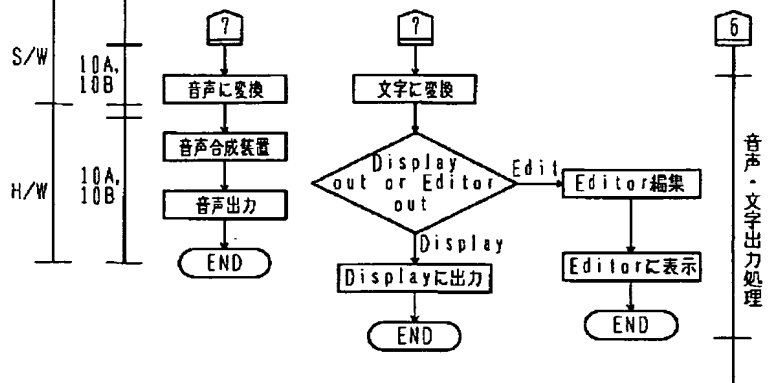
【図4】

処理過程（その2）



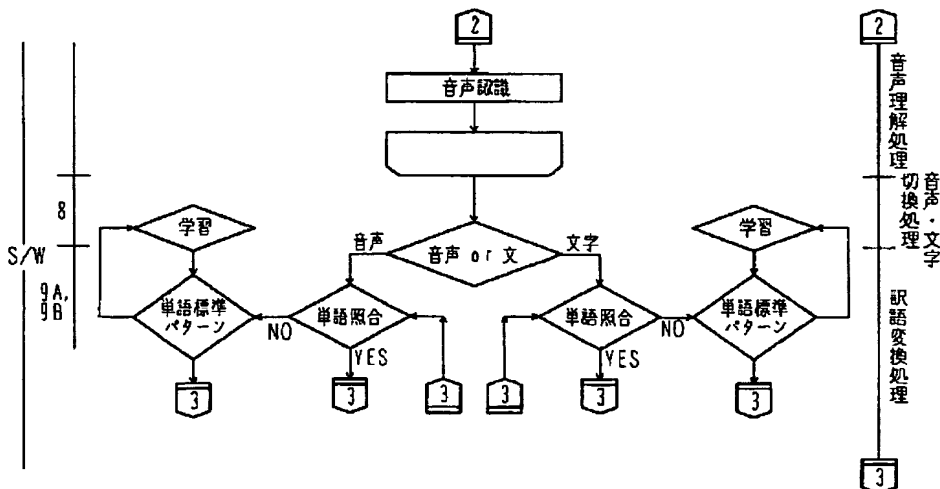
【図9】

処理過程（その7）



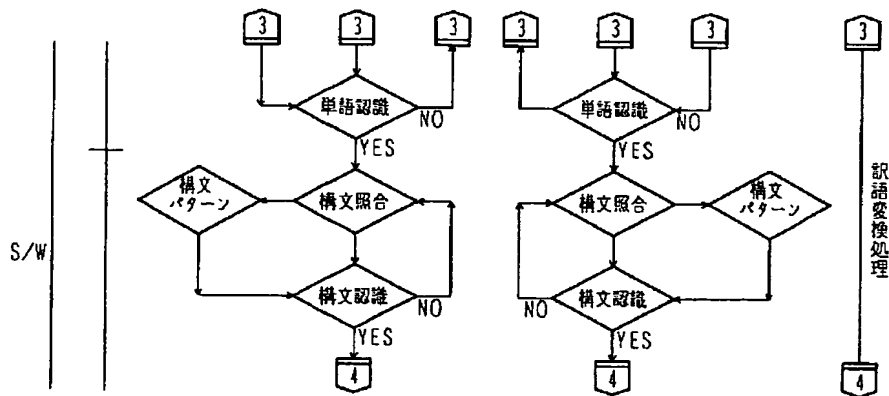
【図5】

処理過程（その3）



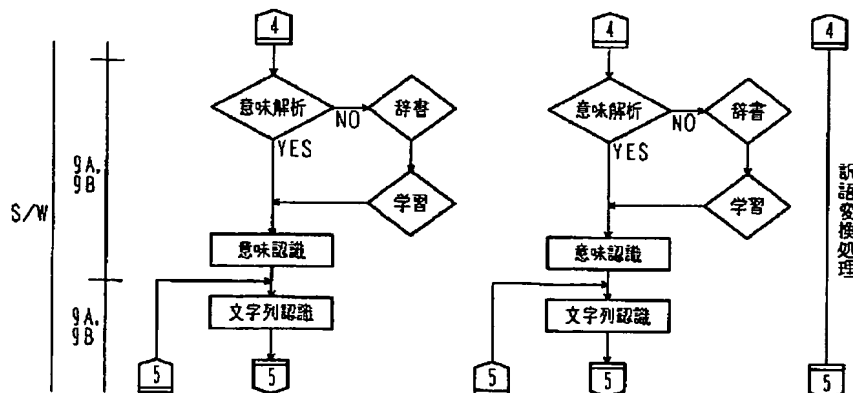
【図6】

## 処理過程（その4）



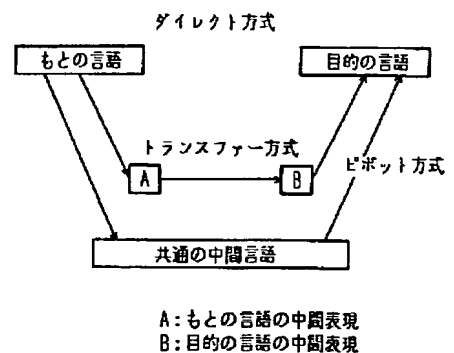
【図7】

## 処理過程（その5）



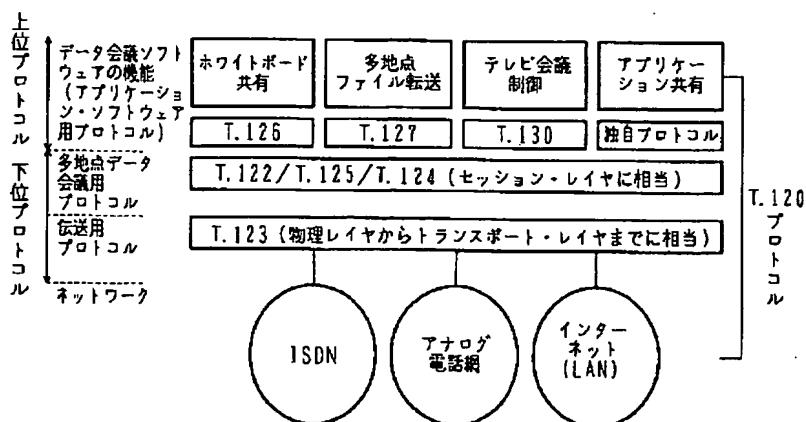
【図13】

## 3つの変換方式の組合せ



【図16】

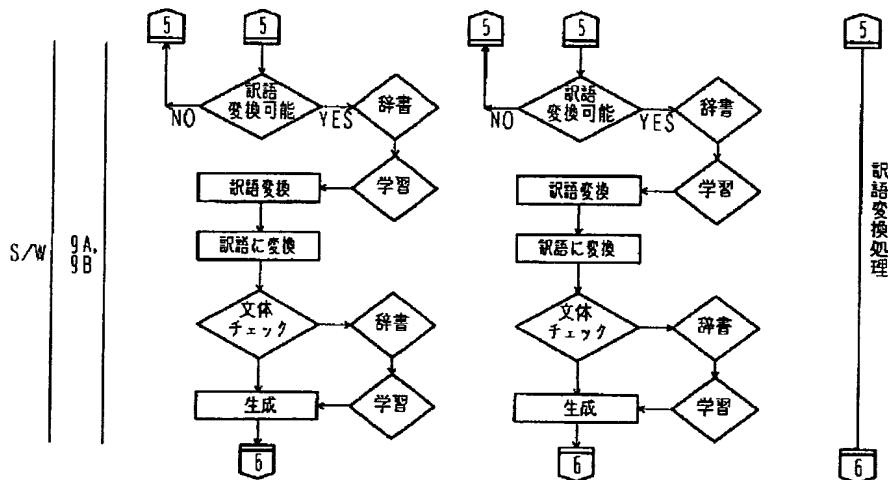
## T.120プロトコルの構成





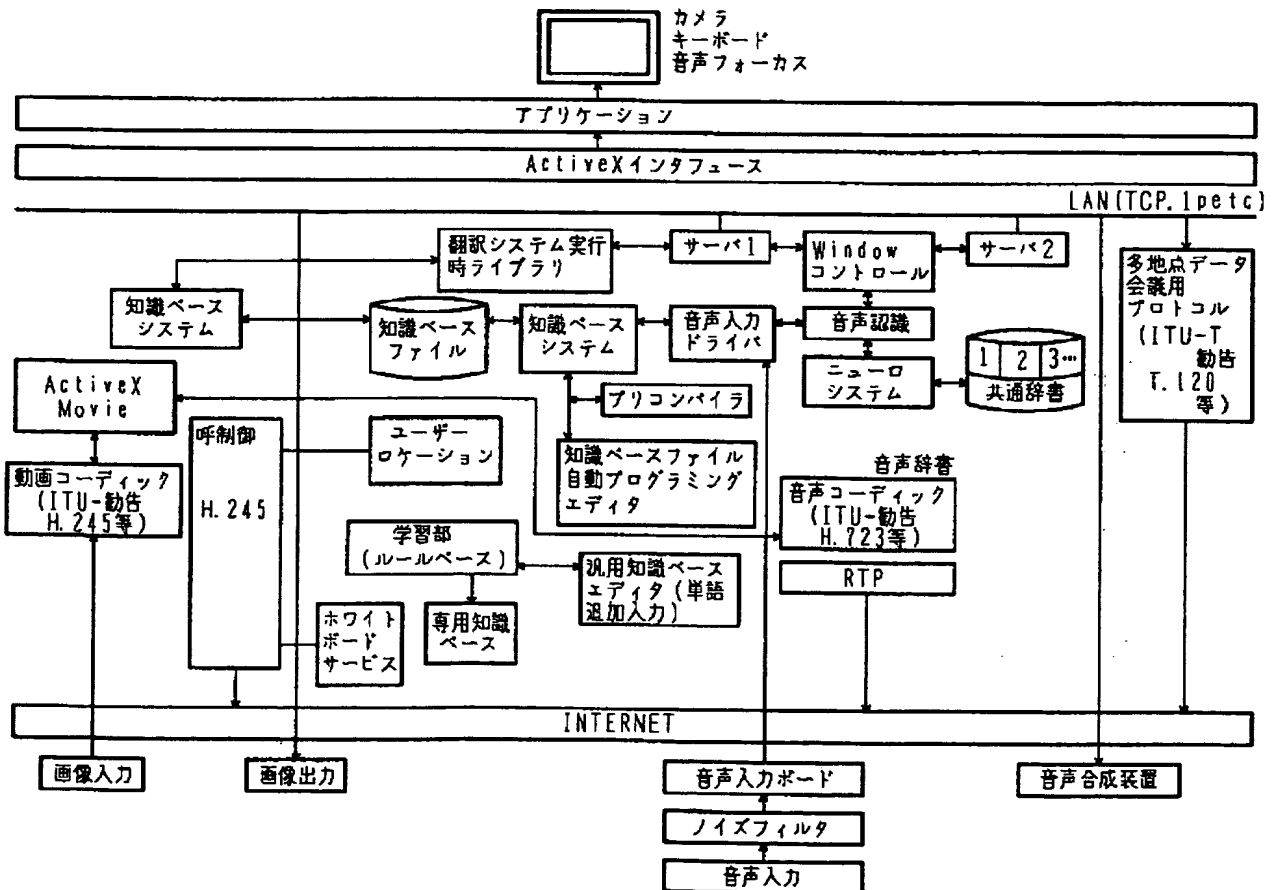
【図8】

処理過程 (その6)



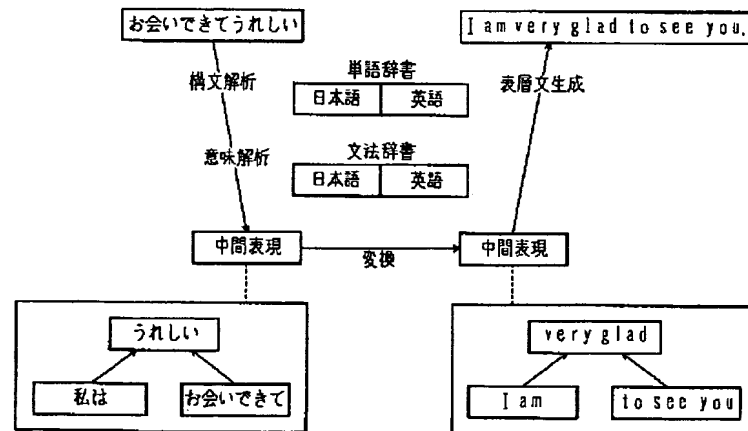
【図10】

訳語機能付デスクトップ会議システムのインタフェース (例)



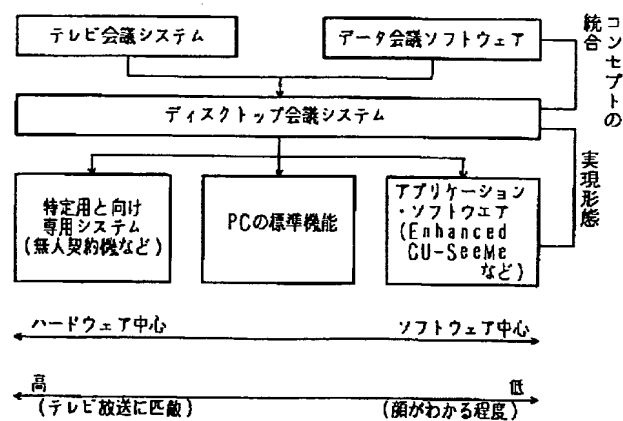
【図12】

トランスファー方式による翻訳システム構成図



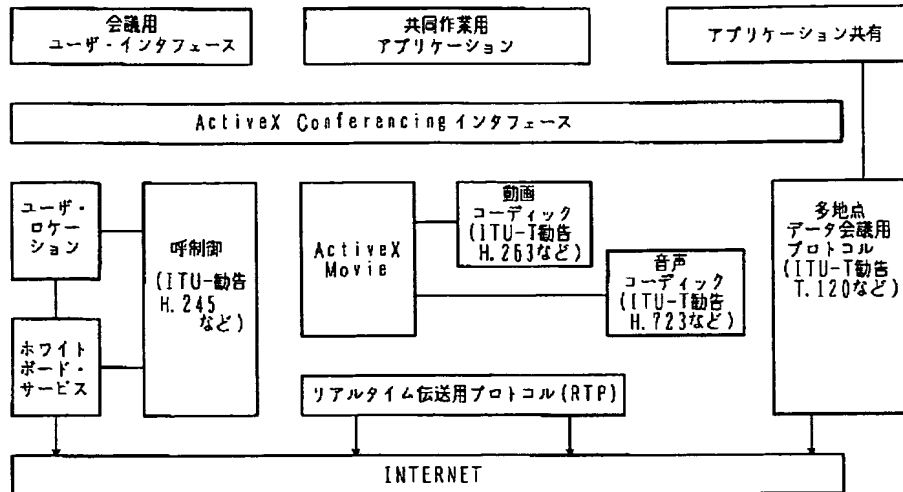
【図14】

TV会議システムの実現形態



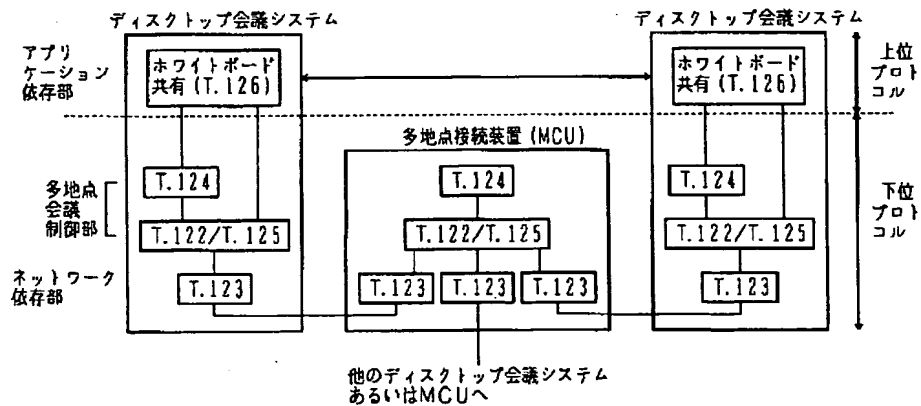
【図15】

## ActiveX Conferencingのアーキテクチャ



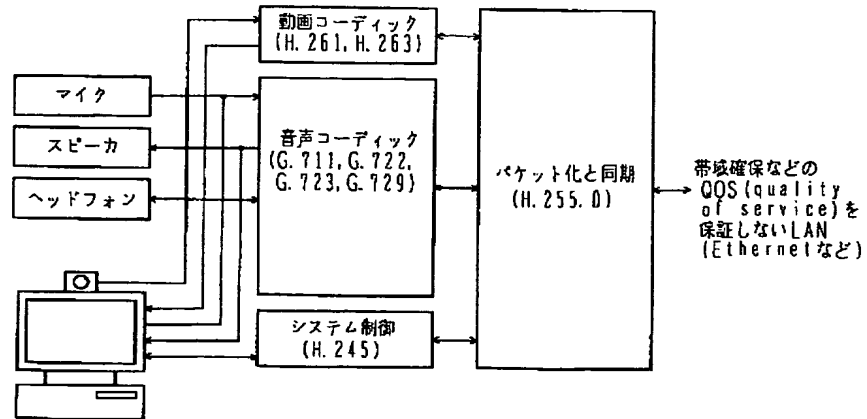
【図17】

## T. 120プロトコルを実装したデスクトップ会議システムとMCU



【図 1 8】

## H. 323 準拠のテレビ会議システムのアーキテクチャ



フロントページの続き

(51) Int. Cl. <sup>6</sup>H 0 4 M 3/42  
3/56

識別記号

F I

H 0 4 M 3/56  
G 0 6 F 15/38C  
V